

RECONOCIMIENTO DE EXPRESIONES FACIALES PARA INTERACCIÓN CON EL COMPUTADOR

Jesús Alfredo Rodríguez Hernández (1), Néstor Darío Duque Méndez (2)

1 [Ing. Mecatrónica, Universidad de Guanajuato - DICIS] | Dirección de correo electrónico: [ralphhlaurent@hotmail.com]

2 [Departamento de Informática y Computación, Universidad Nacional de Colombia – Sede Manizales, Campus La Nubia]

| Dirección de correo electrónico: [ndduqueme@unal.edu.co] (Candara o pts. Color blanco 50%)

Abstract

La interacción Humano-computadora (HCI por sus siglas en inglés) es un campo emergente de la ciencia destinado a proporcionar formas naturales para los humanos para utilizar las computadoras como ayuda. Se dice que para que las computadoras sean capaces de interactuar con los humanos, estas necesitan las habilidades de comunicación de los humanos. Una de estas habilidades es la habilidad de entender el estado emocional de una persona. La manera más expresiva en que los humanos muestran las emociones es a través de las expresiones faciales. En esta investigación se presenta un sistema para el reconocimiento de emociones utilizando la librería de OpenCV. El sistema está basado en Haar Feature-based Cascade Classifier para la detección del rostro y en el algoritmo de Fisherfaces utilizado para el reconocimiento de emociones.

Palabras Clave

Reconocimiento de Emociones; Interacción Humano-computadora ; Haar Feature-based Cascade Classifier; Fisherfaces; OpenCV

Introducción

En años recientes existe un creciente interés en mejorar todos los aspectos de la interacción entre humanos y computadoras. Este campo emergente ha sido un tema de interés de investigación para científicos de diferentes áreas, por ejemplo: ciencias de la computación, ingeniería, psicología, y neurociencia. Estos estudios se enfocan no solo en mejorar la interfaz de la computadora, sino que también en mejorar las acciones que la computadora toma en base a la realimentación del usuario. La realimentación por parte del usuario ha sido tradicionalmente a través del teclado y el ratón. También otros dispositivos han sido desarrollados para interfaces con aplicaciones específicas, tales como los joysticks, trackballs y pantallas táctiles. El rápido avance de la tecnología en recientes años ha hecho que las computadoras sean más baratas y más poderosas, y ha hecho que el uso de micrófonos y cámaras sea asequible y de fácil disponibilidad. Los micrófonos y cámaras permiten a la computadora ver y escuchar, y utilizar esta información para actuar. Un buen ejemplo de esto es el proyecto "Smart - Kiosk" que es realizado por los laboratorios de investigación de Compaq [1]. Para que verdaderamente logremos una interacción efectiva entre Humano-computador (HCI), la computadora necesita ser capaz de interactuar de forma natural con el usuario, similar a la forma en que toma lugar la interacción Humano-humano. Los humanos interactúan entre ellos principalmente a través del habla (lenguaje), pero también a través de gestos del cuerpo, para enfatizar una determinada parte del habla, y para mostrar emociones. Emociones son mostradas por medios visuales, vocales y psicológicos. Existe una creciente cantidad de evidencia que muestra que las habilidades emocionales son parte de lo que llamamos "inteligencia" [2,3].

Hay demasiadas maneras en las cuales los humanos muestran sus emociones. La manera más natural para mostrar las emociones es utilizando expresiones faciales.

Desde principios de los 1970s han existido varios estudios de las expresiones faciales humanas. Ekman y otros [4] encontró evidencia para apoyar la universalidad en expresiones faciales. Estas expresiones faciales universales son aquellas que representan la felicidad, tristeza, enojo, miedo, sorpresa y disgusto. Ellos estudiaron expresiones en muchas culturas, incluyendo las preliterarias, y encontraron atributos comunes en la expresión y reconocimiento de emociones en el rostro. Hay diferencias también: por ejemplo la cultura japonesa suprimirá su expresión facial real en la presencia de autoridades. Los bebés parecen exhibir un gran rango de expresiones faciales sin ser enseñados; esto indica que estas expresiones son innatas.

Este trabajo describe un sistema de reconocimiento de emociones utilizando una webcam como entrada. Nuestro trabajo se enfoca en inicialmente detectar el rostro humano y posteriormente en clasificar la emoción humana mediante el algoritmo de Fisherfaces.

Materiales y métodos

Nuestro sistema consta de 2 partes principales generalizadas: la detección del rostro y la clasificación de estos.



IMAGEN 1: Partes del Sistema

Detección del Rostro

Como nuestro detector de rostro, nosotros elegimos un clasificador rápido y robusto propuesto por Viola y Jones [5] y mejorado por Lienhart y otros [6,7].

Primero, un clasificador es entrenado con cientos de vistas de muestras de un objeto particular (en nuestro caso un rostro), llamados ejemplos

positivos, que son escalados al mismo tamaño (por ejemplo 20x20), y ejemplos negativos -imágenes arbitrarias del mismo tamaño

Después de que el clasificador está entrenado, este puede ser aplicado a una región de interés (del mismo tamaño que se utilizó durante el entrenamiento) en una imagen de entrada. La salida del clasificador entrega un “1” si la región es propensa a mostrar el objeto (el rostro), y “0” si no lo es. Para buscar el objeto en la imagen completa uno puede mover la ventana de búsqueda a través de la imagen y así revisar todos los lugares utilizando el clasificador. El clasificador es diseñado de tal forma que pueda fácilmente redimensionado para ser capaz de encontrar los objetos de interés a diferentes tamaños, lo cual es más eficiente que redimensionar la imagen por si misma. Así, para encontrar el objeto desconocido (el rostro) en la imagen el procedimiento de escaneo debe ser hecho varias veces a diferentes escalas.

La palabra “cascada” en el nombre del clasificador significa que el clasificador resultante consiste de varios clasificadores simples que son aplicados subsecuentemente a una región de interés (ROI) hasta que en una etapa el candidato es rechazado o bien pasa todas las etapas. La palabra “boosted” significa que los clasificadores en cada etapa de la cascada son complejos por si mismos y que ellos están contruidos por clasificadores básicos que utilizan una de las cuatro técnicas de “boosting”. El actual algoritmo utiliza las siguientes características:

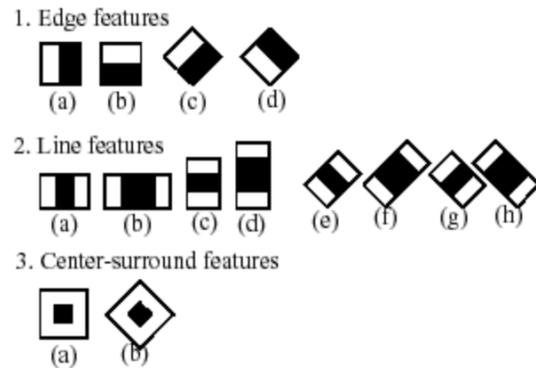


IMAGEN 2: Características Haar-like.

La característica utilizada en un clasificador particular está especificada por su forma (1a, 2b, etc), posición dentro de la región de interés y la escala.

Fisherfaces

El Análisis Discriminante Lineal desempeña una reducción dimensional específica de la clase y fue inventada por el gran estadístico Sir R. A. Fisher. El satisfactoriamente utilizó este método para clasificar flores en 1936. Para encontrar la combinación de características que separa mejor las clases el Análisis Discriminante Lineal maximiza la razón de “dispersión entre clases” a “dispersión dentro de las clases”. La idea es que las mismas clases deben agruparse firmemente cerca, mientras que clases diferentes están tan lejos como es posible de cada una de las otras en la representación de baja dimensión.

RESULTADOS Y DISCUSIÓN

Nuestro sistema trabaja en tiempo real ya que opera sobre cada uno de los cuadros tomados, utiliza un clasificador estático. Con este clasificador se pierde información de los patrones temporales que se presentan en la secuencia de cuadros

(patrones que puede ofrecer un clasificador dinámico), esto quiere decir que nuestro sistema pierde precisión pero al mismo tiempo se gana velocidad de respuesta.

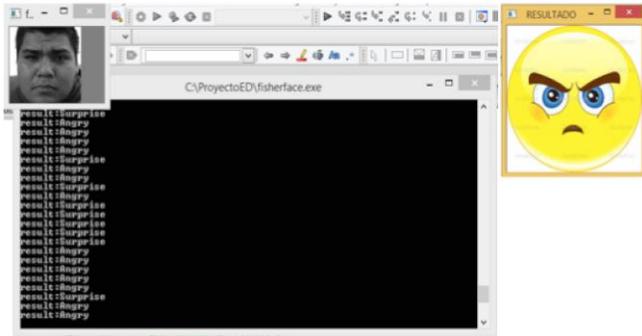


IMAGEN 3: Reconocimiento de rostro feliz

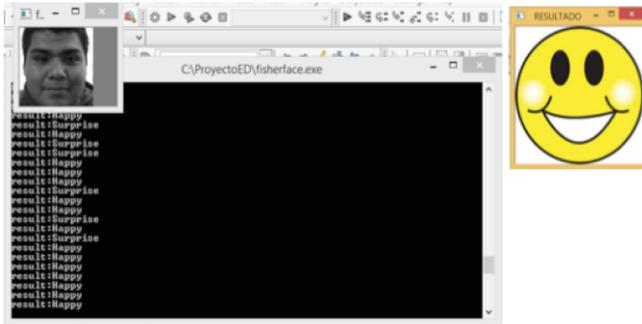


IMAGEN 4: Reconocimiento de rostro feliz

El método utilizado para la detección de rostro es invariante a la escala y muy eficaz aunque es necesario que el usuario se encuentre de frente a la webcam.

El método de Fisherfaces aprende una matriz de transformación específica de la clase, de esta forma no captura iluminación. El análisis Discriminante más bien encuentra las características faciales para discriminar. Es importante mencionar que el desempeño del método Fisherfaces fuertemente depende de la información de entrada.

El sistema responde muy bien a 4 de las 7 emociones consideradas por lo que el sistema trabaja de buena forma pero aun con algunos detalles.

CONCLUSIONES

El sistema trabaja en tiempo real aunque su precisión todavía debe mejorarse, se pueden incluir más características o patrones para esto, tal como patrones temporales, características geométricas o características de textura.

Es un sistema que se puede utilizar para la mejor interacción entre Humano-computador ya que reconoce la parte afectiva.

AGRADECIMIENTOS

Este trabajo lo realicé con la ayuda de mi asesor el investigador Néstor Darío Duque Méndez por lo cual agradezco su apoyo y el de su grupo de trabajo GAIA. También quiero agradecer a mi querida institución la Universidad de Guanajuato y al Programa Delfín por brindarme esta oportunidad de iniciarme en la investigación.

REFERENCIAS

- [1] A. Garg, V. Pavlovic, J. Rehg, and T. S. Huang. Audio–visual speaker detection using dynamic Bayesian networks. In *Proc. of 4rd Intl Conf. Automatic Face and Gesture Rec.*, pages 374–471, 2000.
- [2] P. Salovey and J.D. Mayer. Emotional intelligence. *Imagination, Cognition and Personality*, 9(3):185– 211, 1990.
- [3] D. Goleman. *Emotional Intelligence*. Bantam Books, New York, 1995.
- [4] P. Ekman Strong evidence for universals in facial expressions. *Psychol. Bull.*, 115(2): 268–287, 1994
- [5] P. Viola, M. Jones. Rapid Object Detection Using a Boosted Cascade of Simple Features. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, vol. 1, pp. 511-518, 2001.
- [6] R. Lienhart, J. Maydt. An extended set of haar-like features for rapid object detection. Proceedings of the IEEE International Conference on Image Processing, Rochester, New York, vol. 1, pp. 900-903, 2002.
- [7] R. Lienhart, A. Kuranov, V. Pisarevsky. Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection. Intel Corporation, Technical report, 297–304, 2002.